



## RESEARCH LETTER

10.1002/2017GL074088

## Key Points:

- Seasonal tropical SST forecast skill of operational National Multi-Model Ensemble (NMME) is close to that of simpler linear inverse model
- Since operational SST forecast skill is only slightly lower than the estimated potential skill, we may be near the predictability limit
- In the western Pacific, improvement of operational ensemble is possible since its skill is much lower than the linear inverse model skill

## Supporting Information:

- Supporting Information S1

## Correspondence to:

M. Newman,  
matt.newman@noaa.gov

## Citation:

Newman, M., and P. D. Sardeshmukh (2017), Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures?, *Geophys. Res. Lett.*, *44*, 8520–8529, doi:10.1002/2017GL074088.

Received 8 MAY 2017

Accepted 1 AUG 2017

Accepted article online 3 AUG 2017

Published online 19 AUG 2017

## Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures?

Matthew Newman<sup>1,2</sup>  and Prashant D. Sardeshmukh<sup>1,2</sup><sup>1</sup>CIRES, University of Colorado Boulder, Boulder, Colorado, USA, <sup>2</sup>NOAA/PSD, Boulder, Colorado, USA

**Abstract** The predictability of seasonal anomalies worldwide rests largely on the predictability of tropical sea surface temperature (SST) anomalies. Tropical forecast skill is also a key metric of climate models. We find, however, that despite extensive model development, the tropical SST forecast skill of the operational North American Multi-Model Ensemble (NMME) of eight coupled atmosphere-ocean models remains close both regionally and temporally to that of a vastly simpler linear inverse model (LIM) derived from observed covariances of SST, sea surface height, and wind fields. The LIM clearly captures the essence of the predictable SST dynamics. The NMME and LIM skills also closely track and are only slightly lower than the potential skill estimated using the LIM's forecast signal-to-noise ratios. This suggests that the scope for further skill improvement is small in most regions, except in the western equatorial Pacific where the NMME skill is currently much lower than the LIM skill.

## 1. Introduction

The North American Multi Model Ensemble (NMME) global prediction system [Kirtman *et al.*, 2014] was recently developed to harness the idea that combining forecasts from several state-of-the-art coupled atmosphere-ocean general circulation models (CGCMs) could yield better skill than forecasts using any single CGCM [Hagedorn *et al.*, 2005; Jin *et al.*, 2008; Kirtman and Min, 2009; DelSole *et al.*, 2014]. And indeed the grand NMME mean forecasts are significantly more skillful at predicting El Niño–Southern Oscillation (ENSO)-related sea surface temperature (SST) anomalies than ensemble mean forecasts using the individual models in the ensemble [Becker *et al.*, 2014; Barnston *et al.*, 2015]. The NMME system, used for seasonal predictions since 2011, was made an operational forecast system in 2016.

Given this recent improvement in ENSO prediction, one might wonder if substantial room for further improvement exists (e.g., in model development or initialization techniques) or if seasonal forecast systems are now near an intrinsic predictability limit associated with the chaotic nature of the coupled tropical ocean-atmosphere system. Previous studies in a perfect model framework had suggested that ENSO events might be predictable by some measures up to 2 years ahead [Collins *et al.*, 2002; Chen *et al.*, 2004; Luo *et al.*, 2008; Wittenberg *et al.*, 2014; Larson and Kirtman, 2017]. Some recent studies have been less optimistic in this regard, basing their predictability estimates on multimodel ensembles of retrospective CGCM forecasts (hindcasts) and identifying predictable signals with the ensemble mean forecasts and unpredictable noise with the ensemble spread [Chen *et al.*, 2015; Kumar *et al.*, 2016]. All such predictability estimates are of course model dependent, limited by the CGCMs' ability to realistically represent not only the predictable signals but also the unpredictable noise.

In this study we adopt an alternative empirical approach to estimate tropical Indo-Pacific SST predictability using a linear inverse model [Penland and Sardeshmukh, 1995] (LIM). The LIM is a statistically stationary, stochastically forced, multivariate linear model whose parameters are derivable in principle from the known nonlinear physics of the coupled tropical atmosphere-ocean system [Penland and Sardeshmukh, 1995; Moore and Kleeman, 1999] but can also be estimated more simply from observed covariances of system components as done here. Earlier studies demonstrated that such a LIM captures observed tropical dynamics including the evolution of different types of ENSO events [Capotondi *et al.*, 2013] and ENSO dynamical feedback processes [Newman *et al.*, 2011b], as well as the spectral characteristics of ENSO variability from weekly to multidecadal time scales [Newman *et al.*, 2009; Ault *et al.*, 2013], supporting its use here to benchmark both the actual and potential forecast skill.

Tropical SST predictability studies often focus on specific regional ENSO indices, but comprehensive assessments should be basin wide, given the potentially important pattern and amplitude differences

among individual ENSO events [Capotondi *et al.*, 2013] and also the potential for Indian and western Pacific SST anomalies to have global impacts [Barsugli and Sardeshmukh, 2002]. Therefore, we compare the spatial patterns as well as year-to-year variations of the skills of the LIM, the ensemble mean forecasts of the individual models in the NMME, and the grand NMME mean forecasts over the period 1982–2016. Their similarity to each other and to the potential skill estimated in a perfect LIM framework suggests that the seasonal predictability of tropical Indo-Pacific SST anomalies is effectively linear, apart from minor differences likely due to model deficiencies of the specific LIM and CGCMs utilized here, and that we may now be near the intrinsic limit of tropical Indo-Pacific SST predictability.

## 2. Forecast Models and Metrics

### 2.1. Linear Inverse Model

Penland and Sardeshmukh [1995] showed that many features of observed seasonal tropical SST variability and predictability are captured by a multivariate linear stochastically forced system

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \mathbf{F}_s, \quad (1)$$

where  $\mathbf{x}(t)$  is a state vector of tropical climate anomalies,  $\mathbf{B}$  is a stable linear matrix operator representing feedbacks among the components of  $\mathbf{x}$ , and  $\mathbf{F}_s$  is white noise which may be spatially coherent and whose amplitudes may be seasonally modulated. Such a multivariate model allows for linearly predictable transient anomaly growth as well as decay, unlike a univariate red noise model, and approximates a chaotic dynamical system whose nonlinear processes decorrelate much faster in time than its linear processes [Hasselmann, 1976; Penland, 1996]. These nonlinear processes are represented by the stochastic forcing  $\mathbf{F}_s$  in (1), and their systematic feedbacks on  $\mathbf{x}$ , if any, are absorbed in  $\mathbf{B}$ . Both effects are ignored in traditional linear models of departures of system trajectories from a reference state or a reference trajectory that assume the departures to be small, and the nonlinear effects to be even smaller.

Penland and Sardeshmukh constructed a LIM of tropical SSTs of the form equation (1) by estimating both  $\mathbf{B}$  and the statistics of  $\mathbf{F}_s$  from observations. In principle, a LIM could represent the entire climate anomaly state vector in  $\mathbf{x}$ , but in practice a truncated set of variables is used to (1) focus on the climate subsystem under consideration (here the tropics), (2) minimize redundant information provided by additional closely related variables, and (3) robustly estimate  $\mathbf{B}$  using limited observational data sets. Penland and Sardeshmukh included only SST anomalies (SSTA) in  $\mathbf{x}$ , reasoning that on seasonal time scales the SST-induced changes in surface winds drive an immediate thermocline response, and wave dynamics allow rapid deeper ocean adjustment to the SST changes [Neelin, 1991; Neelin and Jin, 1993]. Subsequent studies [Johnson *et al.*, 2000; Xue *et al.*, 2000; Newman *et al.*, 2011b] showed that explicitly including some oceanic heat content measure in  $\mathbf{x}$  improves the LIM by capturing ENSO recharge-discharge physics [Neelin *et al.*, 1998] and other deeper ocean dynamics not enslaved to SSTA on seasonal time scales. Our approach follows Newman *et al.* [2011b], modifying their  $\mathbf{x}$  by using (1) monthly instead of seasonal anomalies, taken from the observational data sets of the 1958–2010 period listed in Table S1 in the supporting information [Kalnay *et al.*, 1996; Rayner *et al.*, 2003; Ishii *et al.*, 2005; Smith *et al.*, 2008; Balmaseda *et al.*, 2013], (2) sea surface height (SSH) instead of thermocline depth anomalies, given the interest in SSH forecasts for their own sake [Chowdhury and Chu, 2015], and (3) 200 and 850 hPa horizontal winds instead of surface zonal wind stresses to represent the atmospheric ENSO component. The anomaly fields were filtered in each field's empirical orthogonal function (EOF) space defined in the global tropical strip 24°S–24°N, retaining the leading 18/6/4 EOFs representing about 85/63/25% of the domain integrated SST/SSH/wind anomaly variances. The  $28 \times 28$  matrix  $\mathbf{B}$  was then estimated using the 0 and 1 month lag covariance matrices of the 28-component state vector  $\mathbf{x}$ .

$\mathbf{B}$  and hindcast skill were cross-validated by excluding 5 year periods (i.e., 10% of the data) at a time, estimating  $\mathbf{B}$  using the remaining data, and generating hindcasts for the independent 5 year periods. Other key details of LIM construction, including the important “tau test” [Penland and Sardeshmukh, 1995] for the validity of the linear approximation (Figure S1), are described in the supporting information; none of these details affect our conclusions. Note that the LIM used here and its dynamics are consistent with other related studies [Newman *et al.*, 2011b; Vimont *et al.*, 2014; Capotondi and Sardeshmukh, 2015].

## 2.2. NMME Hindcasts

The LIM's hindcast skill was compared to that of the NMME models used operationally by National Centers for Environmental Prediction (NCEP) (Table S2); including models from earlier stages of the NMME project did not alter our conclusions. All CGCM forecast models drift toward their own climate at long forecast lead times, causing a forecast bias. We removed each model's bias from each of its ensemble members as a function of both calendar month and forecast lead time, using the 1982–2010 hindcast data set and the LIM's cross-validation intervals. Following *Barnston et al.* [2015], the grand multimodel ensemble mean (NMME mean) forecasts were then determined using such bias-corrected ensemble members of all the models. The bias correction improves the skill of each model and of the NMME mean [*Barnston et al.*, 2015].

The NMME forecasts were initialized on or near (for staggered starts) the first day of each month. The "Month 0.5" forecast was then the mean of the first month of the forecast run, i.e., centered in the middle of the calendar month. The equivalent LIM forecast was initialized with the monthly mean observations centered on the previous month, so that the 1 month lead LIM forecast and the NMME Month 0.5 forecast were verified at the same time. We renamed both these forecasts the "Month 1" forecast and so on for increasing forecast leads.

## 2.3. Evaluating Forecast Skill and Predictability

We used two skill measures: anomaly correlation (AC) and root-mean-square error skill score (RMS skill score) defined as  $\varepsilon \equiv 1 - \hat{\sigma}$ , where  $\hat{\sigma} = \sigma / \sigma_{\text{obs}}$ ,  $\sigma$  is the RMS forecast error, and  $\sigma_{\text{obs}}$  is the observed climatological RMS value. The overall skill over the tropical Indo-Pacific domain was assessed using the standardized RMS error and pattern correlation between the forecast and verification SSTA fields in the region 24°S–24°N, 30°E–60°W.

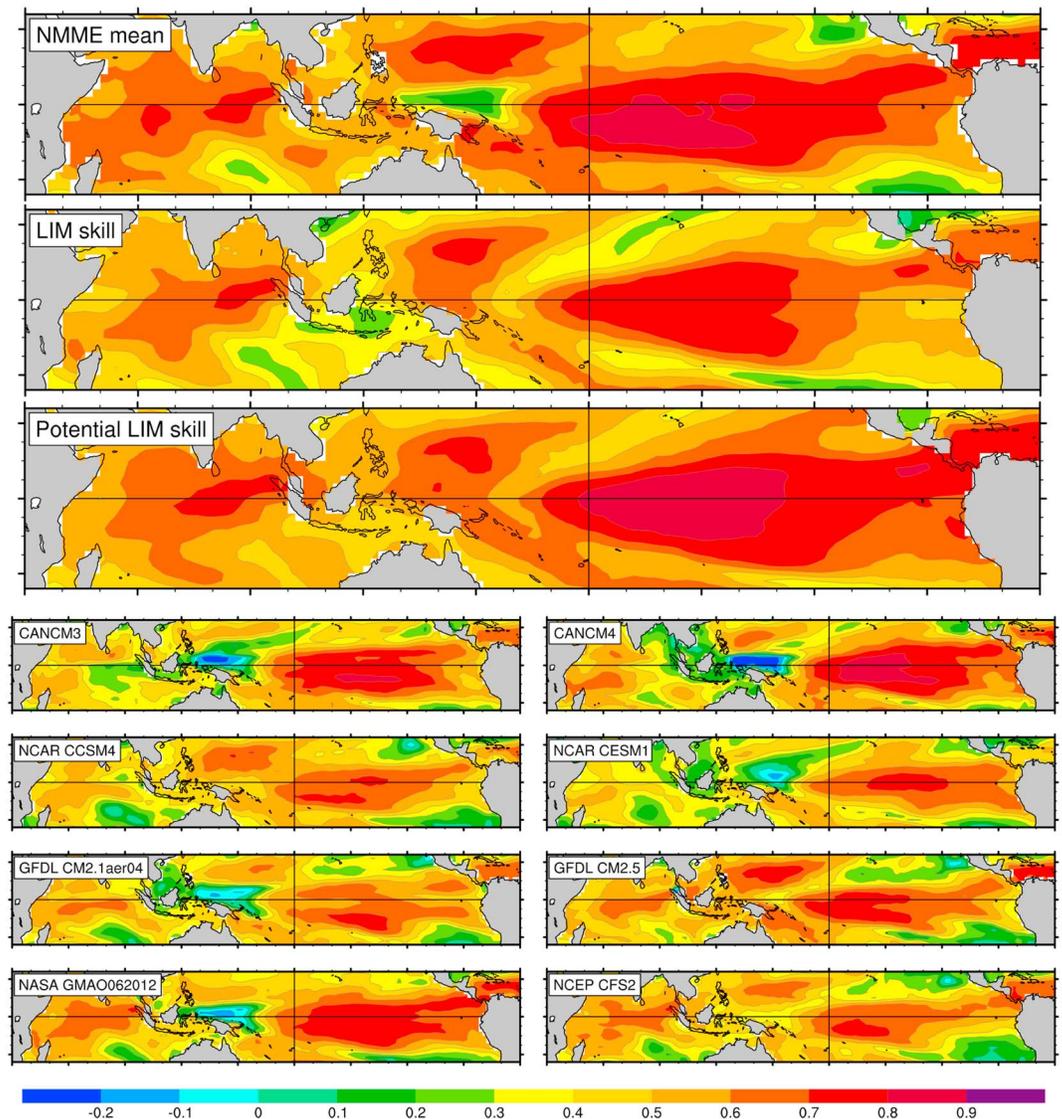
Potential skill in a chaotic system may be defined as the skill of a perfect model with infinitesimal errors in initial conditions. For infinite-member ensemble perfect model forecasts, it can be shown that the average AC between forecasts at lead time  $\tau$  and the corresponding verification is a function of the forecast signal-to-noise ratio [*Sardeshmukh et al.*, 2000]. The distinction between the predictable forecast signal and unpredictable noise is explicit in (1) and yields explicit expressions for the predictable forecast signal covariance and forecast error covariance matrices which can be used to determine the potential LIM skill in terms of both AC skill [*Newman et al.*, 2003] and RMS skill scores (see supporting information).

## 3. Tropical Indo-Pacific SST Forecast Skill and Predictability

Figure 1 shows Month 6 local AC skill of the LIM forecasts, each individual NMME model's ensemble mean forecasts, and the NMME mean forecasts. The pattern and amplitude of the NMME mean and LIM skills are very similar, with equatorial skill maxima east of the dateline, in the central Indian Ocean, and in a broad region east of the Philippines. Skill minima are found in a V-shaped region extending northeastward and southeastward from the equator at about 160°E and in the southeast Indian Ocean. The LIM and NMME mean skills are closer to each other than to any of the individual NMME models, for all forecast lead times (not shown). Note that this similar skill occurs despite initializing the LIM only within its limited EOF space.

Overall, the potential LIM skill (also in Figure 1) has a very similar pattern to that of the actual NMME and LIM skill and only slightly higher values. The highest potential skill is not in the eastern equatorial Pacific, where both the SST forecast signals and noise are large, but rather in the central Pacific where large forecast signals are accompanied by much weaker noise. Not surprisingly, the potential skill is smallest along nodal lines of the dominant ENSO pattern where the forecast signals are generally weakest but is larger again farther westward associated with stronger ENSO-related signals over the Indian Ocean where the noise is also weaker.

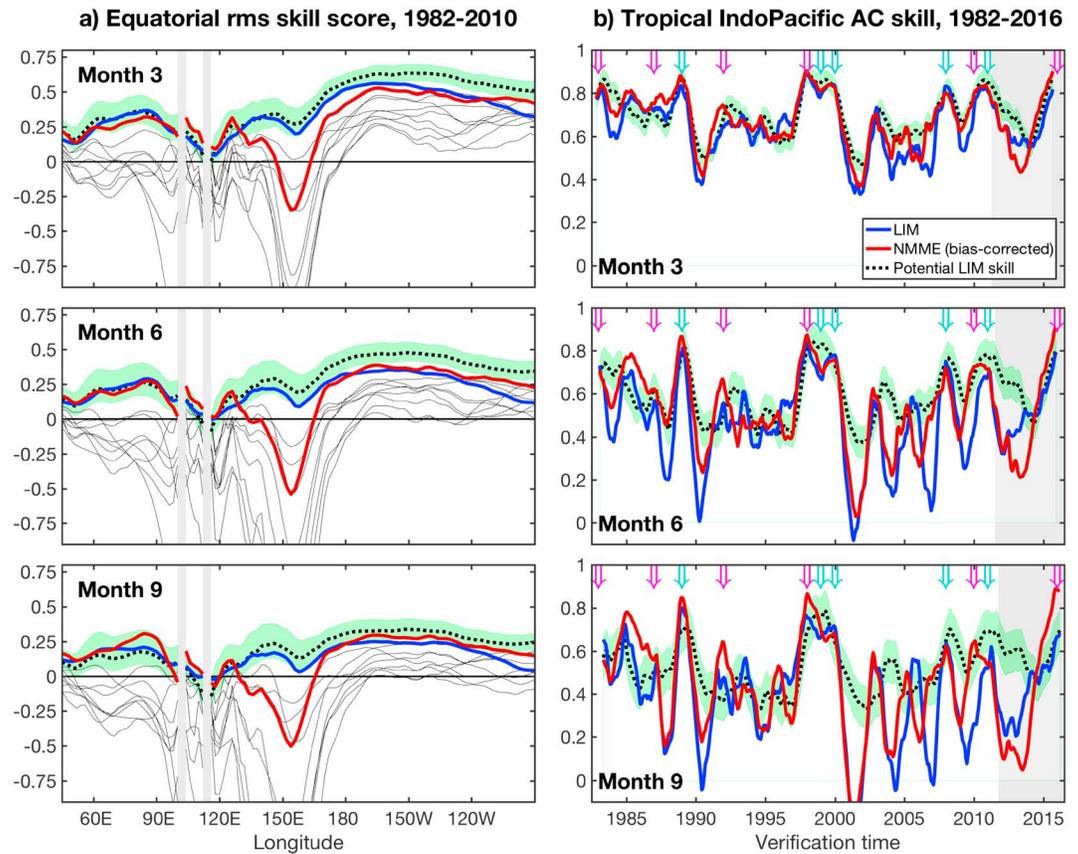
Along the equator, LIM and NMME mean RMS skill scores vary similarly with longitude and forecast lead time (Figure 2a). The LIM and NMME mean have nearly equal skill in the central Pacific and Indian Ocean. The LIM has less skill in the far eastern Pacific and far more skill in the western Pacific where the NMME mean skill (and the skill of all its constituent models; cf. Figure 1) is *negative*. The potential RMS skill score along the equator (dotted line in Figure 2a) closely tracks the LIM and NMME mean hindcast skill. These potential skill estimates are derived from a LIM trained on a limited observational record and so are uncertain to some extent. To estimate their sampling uncertainties, we generated 26,500 years of synthetic data using (1) [*Penland and Matrosova*, 1994; *Newman et al.*, 2011a] and estimated the LIM parameters and associated potential skill using each of its five hundred 53 year segments to determine 90% confidence intervals for the potential skill, shown as the green band in Figure 2a. (See supporting information for details and Figure S2 for the



**Figure 1.** Local anomaly correlation (AC) skill of Month 6 SST anomaly (SSTA) hindcasts over the period 1982–2010. The top three panels show the NMME mean, LIM, and potential LIM skill. The skill of each individual NMME model’s ensemble mean hindcasts is shown in the bottom eight panels. Note that each NMME model hindcast is separately bias-corrected to remove its mean monthly hindcast error, and the bias-corrected forecasts are then averaged to form the NMME mean hindcasts. The LIM is constructed with data for the 1958–2010 period; however, only the cross-validated hindcast skill over the 1982–2010 period is shown. Skill for the full period is in Figure S3a.

confidence intervals of the skill estimates in Figure 1). The LIM’s hindcast skill is within this confidence interval west of the dateline but not in the far eastern Pacific where it is significantly lower than both the potential skill and the NMME mean skill.

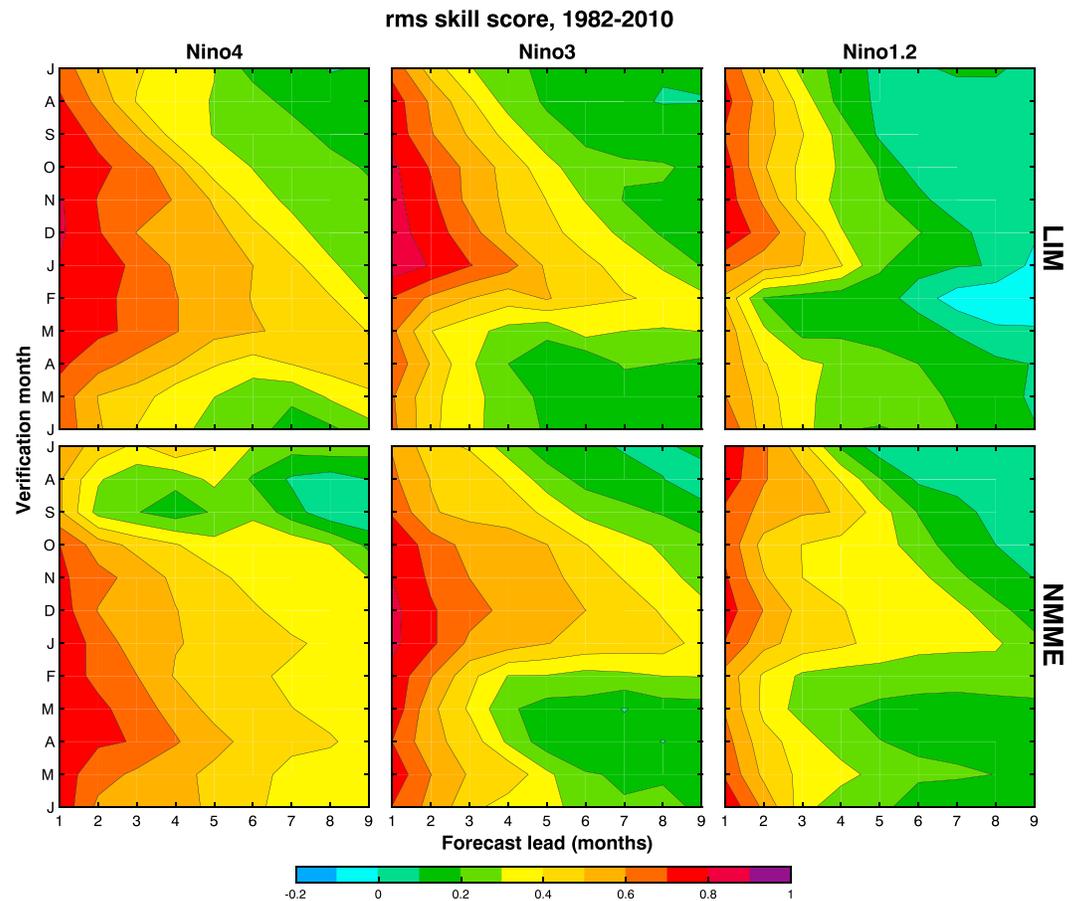
The LIM and the NMME mean also have very similar year-to-year skill variations, and again both are similar to the potential LIM skill variations. Figure 2b shows the 13 month running mean pattern correlations between the predicted and observed tropical Indo-Pacific SSTA fields for lead times of 3, 6, and 9 months. The LIM and NMME mean skill series are highly correlated for all these leads. The potential skill is largely realized by both forecast systems at Month 3, to a lesser extent at Month 6, but poorly at Month 9 after 2000. Overall, the NMME mean skill is slightly higher than the LIM skill. Note also that in a few cases, both the NMME mean and LIM skills are higher than the potential skill. This is not a contradiction. The potential skill represents an *expected* upper skill bound, but actual forecast skill in individual forecast cases can be higher by chance depending on the magnitude of the actually realized unpredictable noise in those cases.



**Figure 2.** Comparison of LIM skill (blue), potential LIM skill (black dots), and the NMME mean skill (red), for forecast leads of 3, 6, and 9 months. Light green shading shows 90% confidence intervals for potential LIM skill (see text for details). (a) Root-mean-square (RMS) skill score of SST anomaly hindcasts averaged between 3°S and 3°N, over the period 1982–2010. Note that since the RMS skill score is defined as 1 minus the standardized RMS error, higher values indicate higher skill. Therefore, forecasts with perfect skill have a score = 1, climatological (zero anomaly) forecasts have score = 0, and forecasts worse than climatology have negative skill. Gray vertical bars indicate land (no SST values). The thin lines show the skill of each individual NMME model’s ensemble mean hindcasts. (b) Year-to-year variations of model hindcast/forecast skill in the tropical Indo-Pacific over the 1980–2016 period. The pattern anomaly correlation (AC) skill between the forecast and verification monthly SST anomalies is determined in the tropical Indo-Pacific domain (24°S–24°N, 30°E–60°W) and then smoothed with a 13 month running mean. The gray vertical shaded bar in each panel shows the “forecast” period (i.e., subsequent to LIM cross-validation and NMME hindcast bias correction periods). Magenta (green) arrows along the top indicate years of >1 sigma warm (< -1 sigma cold) Niño3.4 SST anomalies. The LIM and NMME mean forecast skills are correlated at 0.9, 0.8, and 0.8 for Months 3, 6, and 9, respectively, whereas the LIM and potential LIM skills are correlated at 0.9, 0.7, and 0.6 for months 3, 6, and 9, respectively.

There is a hint of lower skill for several years following the 1997–1998 ENSO event [Barnston *et al.*, 2012], which some studies have suggested as possibly due to an altered tropical base state in a changing climate, with reduced ENSO predictability [Zhao *et al.*, 2016]. Note, however, that since our LIM’s dynamical operator **B** is fixed, all variations of *potential* LIM skill result only from random variations in initial conditions and consequent variations in predictable anomaly growth. Given the correspondence between potential skill and model skill, as well as the similarity between LIM and NMME mean skill, the realized skill variations in Figure 2b are therefore likely due to differing noise realizations as opposed to base state changes, consistent with other recent model and forecast studies [Wittenberg *et al.*, 2014; Kumar *et al.*, 2015; Lee *et al.*, 2016].

It is striking that our **B** matrix, despite having no seasonal variation, captures the essence of the seasonal variation of the NMME mean skill, shown in Figure 3 for three commonly used ENSO indices. Note that for most lead times, the LIM and NMME mean skills vary more with verification month than initialization month. For example, Niño3 and Niño4 forecasts for the summer months are generally much less skillful than for the winter months. The LIM also captures the substantially reduced eastern Pacific (Niño1.2 and Niño3) skill, at

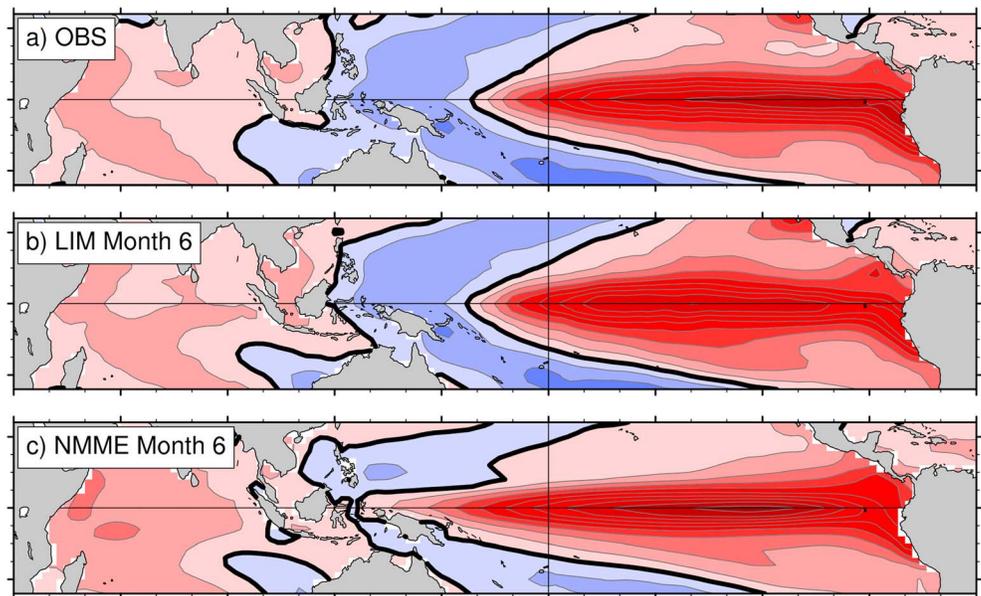


**Figure 3.** Root-mean-square (RMS) skill scores of the LIM and NMME mean hindcasts of three equatorial Pacific SST anomaly indices over the 1982–2010 period. (top row) The LIM skill and (bottom row) the NMME mean skill of area-averaged SST anomalies in the (left column) Nino4, (middle column) Nino3, and (right column) Nino1.2 regions. All panels show skill as a function of forecast lead (abscissa) and verification month (ordinate). For example, skill at the position (4, S) refers to Month 4 hindcasts verifying in September, which are initialized from the previous May monthly mean anomaly for the LIM, and 1 June (or within a few days, depending on the model) for the NMME mean. Note that since the LIM is, by construction, independent of the annual cycle, any seasonal variations in its skill represent either random sampling or seasonal variation in the noise forcing in (1).

any given lead time, of forecasts for spring compared to forecasts for winter, sometimes referred to as the springtime predictability barrier [e.g., *Levine and McPhaden, 2015*].

The NMME's poor skill just west of the dateline (Figure 2a) contributes to its poorer Niño4 skill (Figure 3). The basic reason for this poor skill is that the NMME mean predicted pattern of eastern Pacific warming during El Niño events (and cooling during La Niña events) extends too far west of the dateline compared to observations. This is confirmed by comparing the dominant pattern (leading EOF) of the observed SSTA with that in the Month 6 LIM and NMME mean SSTA hindcasts (Figure 4). While a typical ENSO pattern in a LIM forecast matches observations, with SSTAs of opposite signs in the western and eastern Pacific, the NMME mean typically predicts an ENSO anomaly with the same sign across the entire equatorial Pacific. This pattern is particularly dominant for the individual NMME models with the lowest western Pacific SSTA skill (not shown, but see Figure S3a), which also generally have the highest central Pacific SSTA skill. While the erroneous westward extension of ENSO variability is common in long climate model simulations [e.g., *Joseph and Nigam, 2006; Li and Xie, 2014*], it is notable that this error develops rapidly in 1–2 months from initial conditions in the NMME forecasts (Figure S3c) but not in the LIM forecasts (Figure S3b).

Finally, we depict in Figure 5 on a Taylor diagram [Taylor, 2001] the overall decrease of skill with increasing forecast lead time as color-coded curves for the LIM, NMME mean, and individual NMME model hindcasts.



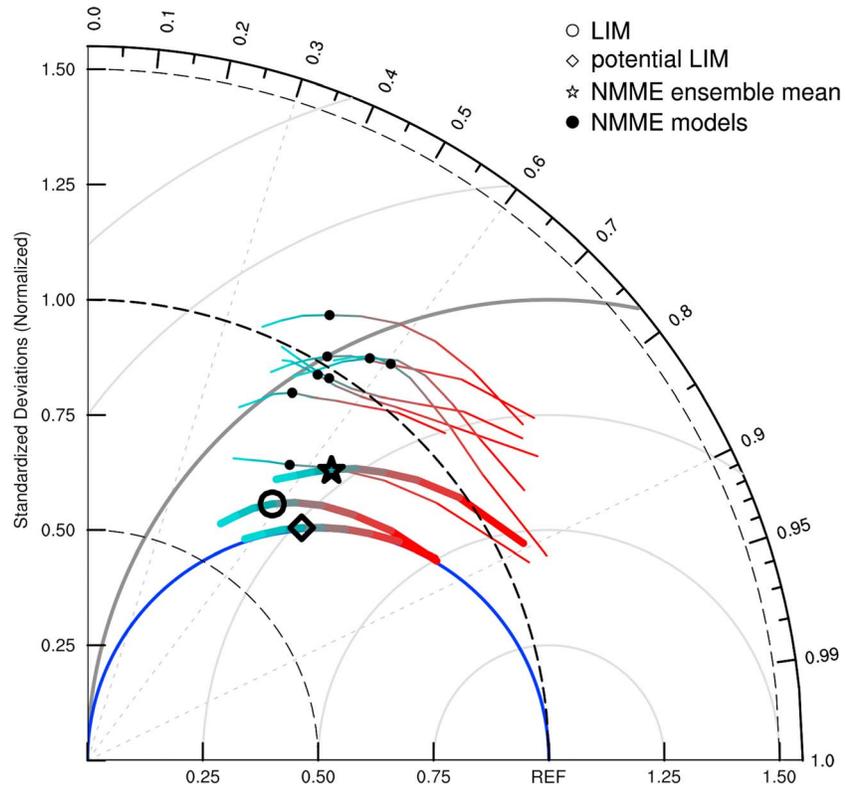
**Figure 4.** The dominant pattern (leading empirical orthogonal function, or EOF) of monthly SST variability in observations and forecasts during 1982–2010. (a) Observed, (b) Month 6 LIM forecasts, and (c) Month 6 bias-corrected NMME mean forecasts. These patterns explain 40%, 62%, and 60% of the variability of their respective fields.

The angular coordinate of the plotted values for each model represents the average pattern correlation of the hindcast SSTA fields with the verification fields over the Indo-Pacific domain, and the polar coordinate (distance from the origin) represents the RMS magnitude of the hindcast fields standardized by that of the verification fields. The plotted point for the verification fields is, by definition, the reference point (REF). The distance of the model points from REF is the standardized RMS forecast error. With increasing forecast lead time, the skill trajectory of all model hindcasts recedes monotonically from REF.

It can be shown (see supporting information) that the skill trajectory of any perfect model whose forecast errors are on average orthogonal to its forecast signals should follow a universal semicircular trajectory on a Taylor diagram (the blue curve in Figure 5), as is indeed true for the potential LIM skill trajectory. The actual LIM hindcast LIM skill is very close to this curve at short lead times but diverges from it at longer lead times. The NMME mean skill roughly parallels the LIM skill, with slightly higher AC and larger forecast RMS magnitudes, which indeed exceed even the observed magnitudes [Barnston *et al.*, 2015] for leads of a few months. The net effect is that the RMS forecast error of the NMME mean is slightly better (~3–4%) than the LIM, but both are larger than the potential LIM error.

#### 4. Discussion

The similarity of the NMME mean skill to the LIM skill could conceivably be due to NMME model misrepresentations of predictable nonlinear dynamics. Some model errors that are plain in long climate runs have persisted through several generations of CGCMs, along with continuing difficulties in capturing the correct balance of ENSO processes [Bellenger *et al.*, 2014]. Averaging across multiple imperfect CGCMs, as in the NMME mean forecasts, may wash out a possibly large nonlinear part of the forecast signals on which there is model disagreement and emphasize the linear part, yielding skill similar to the LIM skill. However, the fact that both the LIM and the NMME mean outperform all the individual models in the NMME ensemble would then suggest that such predictable nonlinear signals are not captured by any of the individual models, which seems unlikely given extensive CGCM developments over the past several decades. A much simpler interpretation of our results is that the nonlinear interactions are not necessarily weak but are essentially unpredictable (as is indeed assumed by the LIM) and cause random forecast errors that are averaged out better in the NMME mean forecasts than in the individual-model ensemble mean forecasts. Still, if future



**Figure 5.** Evolution of SST anomaly skill for Months 1 to 9 lead times. In this Taylor diagram [Taylor, 2001] in polar  $(r, \theta)$  coordinates, the polar coordinate  $r$  (distance from the origin) of a point represents the domain- and time-averaged RMS magnitude of forecast fields normalized by that of the observed fields, the angular coordinate  $\theta$  (measured counterclockwise from the x axis) represents the inverse cosine of the pattern anomaly correlation  $\rho = \cos \theta$  of the forecast and observed fields, and the distance of  $(r, \theta)$  from the reference point REF at  $(1, 0)$  represents the RMS error of the forecast fields normalized by the RMS magnitude of the observed fields. The analysis is performed over the same Indo-Pacific domain used in Figure 2b, for the NMME mean, LIM, and perfect LIM forecasts (thick color-coded curves), as well as for the ensemble mean forecasts of the individual NMME models (thin color-coded curves). Each curve shows the (counterclockwise) decline of forecast skill from Month 1 (red) to Month 9 (aqua). The symbol marker on each curve, placed at the position corresponding to Month 6 skill, also identifies the type of forecast model, shown in the top right corner. The individual NMME models are not named. The blue semicircle shows the trajectory of perfect LIM skill, which also declines counterclockwise for increasing forecast lead. See text for more details.

CGCM improvements do lead to a better consensus on the predictable nonlinear signals, then the NMME mean skill could exceed LIM skill in cases in which those signals are large, possibly in the eastern equatorial Pacific.

On the other hand, our LIM could also be improved. As an empirical model, it is limited by the availability of adequate training data, particularly of SSH and related measures of the ocean state. The LIM could also account better for seasonality, especially in the far eastern Pacific [Mitchell and Wallace, 1992], although the available observational record is likely too short for this [Johnson et al., 2000; Newman et al., 2009], and the LIM already largely captures seasonal variation of NMME mean skill. Including additional variables in the LIM to improve its representation of coupled tropical dynamics poses similar data challenges. It is also important to remember that our particular LIM is derived from a single 53 year observed segment of an essentially stochastically driven multivariate linear process whose parameters might require hundreds of years of data to be determined with sufficient accuracy [Wittenberg, 2009], with consequent uncertainty in the estimated potential skill as shown in Figure 2.

But these issues do not diminish our central result: two fundamentally different forecast systems, one a low-dimensional empirical linear inverse model and the other a multimodel ensemble of comprehensive high-dimensional nonlinear coupled GCMs, have very similar spatially and temporally varying tropical

Indo-Pacific SST skill. This is strong evidence that the predictable dynamics of tropical Indo-Pacific SST variations are essentially linear, and a linear model can therefore be used to quantify that predictability. When used for this purpose, our LIM predicts upper bounds on the spatial/temporal skill variations of both the NMME and LIM forecast systems as a direct consequence of the spatial/temporal variations of its forecast signal-to-noise ratio, that is, of its (effectively infinite-member) ensemble mean forecasts and ensemble spread. The fact that the NMME mean skill and the LIM skill are consistently just below these upper bounds suggests that we may be near the limit of tropical Indo-Pacific SST predictability in most regions, except in the western equatorial Pacific where the NMME mean skill is currently much lower than the LIM skill. One might thus reasonably expect the impacts of future improvements in seasonal tropical Indo-Pacific SST forecast skill to be small, unless the current NMME model error in the western Pacific has a large negative impact on global seasonal predictions.

### Acknowledgments

We acknowledge the agencies that support the NMME-Phase II system, and we thank the participating climate modeling groups (Environment Canada, NASA, NCAR, NOAA/GFDL, NOAA/NCEP, and University of Miami) for producing and making their model output available. NOAA/NCEP, NOAA/CTB, and NOAA/CPO jointly provided coordinating support and led development of the NMME-Phase II system. All NMME hindcast data used in this paper are available from the IRI/LDEO Data Library, <https://iridl.ldeo.columbia.edu>. LIM hindcasts will be made available from NOAA/ESRL/PSD, <https://www.esrl.noaa.gov/psd>. This work was supported by NOAA/CPO-MAPP and NSF AGS 1463643.

### References

- Ault, T. R., C. Deser, M. Newman, and J. Emile-Geay (2013), Characterizing decadal to centennial variability in the equatorial Pacific during the last millennium, *Geophys. Res. Lett.*, *40*, 3450–3456, doi:10.1002/grl.50647.
- Balmaseda, M. A., K. Mogensen, and A. T. Weaver (2013), Evaluation of the ECMWF ocean reanalysis system ORAS4, *Q. J. R. Meteorol. Soc.*, *139*(674), 1132–1161.
- Barnston, A. G., M. K. Tippett, M. L. L'Heureux, S. Li, and D. G. DeWitt (2012), Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing?, *Bull. Am. Meteorol. Soc.*, *93*(5), 631–651.
- Barnston, A. G., M. K. Tippett, H. M. van den Dool, and D. A. Unger (2015), Toward an improved multimodel ENSO prediction, *J. Appl. Meteor. Climatol.*, *54*(7), 1579–1595.
- Barsugli, J. J., and P. D. Sardeshmukh (2002), Global atmospheric sensitivity to tropical SST anomalies throughout the Indo-Pacific Basin, *J. Clim.*, *15*(23), 3427–3442.
- Becker, E., H. van den Dool, and Q. Zhang (2014), Predictability and forecast skill in NMME, *J. Clim.*, *27*(15), 5891–5906.
- Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard (2014), ENSO representation in climate models: From CMIP3 to CMIP5, *Clim. Dyn.*, *42*(7–8), 1999–2018.
- Capotondi, A., and P. D. Sardeshmukh (2015), Optimal precursors of different types of ENSO events, *Geophys. Res. Lett.*, *42*, 9952–9960, doi:10.1002/2015GL066171.
- Capotondi, A., et al. (2013), Understanding ENSO diversity, *Bull. Am. Meteorol. Soc.*, *96*(6), 921–938.
- Chen, D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang (2004), Predictability of El Niño over the past 148 years, *Nature*, *428*(6984), 733–736.
- Chen, M., A. Kumar, and W. Wang (2015), A study of the predictability of sea surface temperature over the tropics, *Clim. Dyn.*, *44*(7–8), 1767–1776.
- Chowdhury, M. R., and P. S. Chu (2015), Sea level forecasts and early-warning application: Expanding cooperation in the South Pacific, *Bull. Am. Meteorol. Soc.*, *96*(3), 381–386.
- Collins, M., D. Frame, B. Sinha, and C. Wilson (2002), How far ahead could we predict El Niño?, *Geophys. Res. Lett.*, *29*(10), 1492, doi:10.1029/2001GL013919.
- DelSole, T., J. Nattala, and M. K. Tippett (2014), Skill improvement from increased ensemble size and model diversity, *Geophys. Res. Lett.*, *41*, 7331–7342, doi:10.1002/2014GL060133.
- Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer (2005), The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept, *Tellus A*, *57*(3), 219–233.
- Hasselmann, K. (1976), Stochastic climate models, Part I. Theory, *Tellus*, *28*(6), 473–485.
- Ishii, M., A. Shouji, S. Sugimoto, and T. Matsumoto (2005), Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection, *Int. J. Climatol.*, *25*(7), 865–879.
- Jin, E. K., et al. (2008), Current status of ENSO prediction skill in coupled ocean–atmosphere models, *Clim. Dyn.*, *31*(6), 647–664.
- Johnson, S. D., D. S. Battisti, E. S. Sarachik, S. D. Johnson, and E. S. Sarachik (2000), Empirically derived Markov models and prediction of tropical Pacific sea surface temperature anomalies\*, *J. Clim.*, *13*(1), 3–17.
- Joseph, R., and S. Nigam (2006), ENSO evolution and teleconnections in IPCC's twentieth-century climate simulations: Realistic representation? *J. Clim.*, *19*, 4360–4377.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–471.
- Kirtman, B. P., and D. Min (2009), Multimodel ensemble ENSO prediction with CCSM and CFS, *Mon. Weather Rev.*, *137*(9), 2908–2930.
- Kirtman, B. P., et al. (2014), The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction, *Bull. Am. Meteorol. Soc.*, *95*(4), 585–601.
- Kumar, A., M. Chen, Y. Xue, and D. Behringer (2015), An analysis of the temporal evolution of ENSO prediction skill in the context of the equatorial Pacific Ocean observing system, *Mon. Weather Rev.*, *143*(8), 3204–3213.
- Kumar, A., Z.-Z. Hu, B. Jha, and P. Peng (2016), Estimating ENSO predictability based on multi-model hindcasts, *Clim. Dyn.*, *48*, 39–51.
- Larson, S. M., and B. P. Kirtman (2017), Drivers of coupled model ENSO error dynamics and the spring predictability barrier, *Clim. Dyn.*, *48*(11–12), 3631–3644.
- Lee, J.-W., S.-W. Yeh, and H.-S. Jo (2016), Weather noise leading to El Niño diversity in an ocean general circulation model, *Clim. Dyn.*, 1–13.
- Levine, A. F. Z., and M. J. McPhaden (2015), The annual cycle in ENSO growth rate as a cause of the spring predictability barrier, *Geophys. Res. Lett.*, *42*, 5034–5041, doi:10.1002/2014GL060133.
- Li, G., and S.-P. Xie (2014), Tropical biases in CMIP5 multimodel ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems, *J. Clim.*, *27*, 1765–1780.
- Luo, J.-J., S. Masson, S. K. Behera, and T. Yamagata (2008), Extended ENSO predictions using a fully coupled ocean–atmosphere model, *J. Clim.*, *21*(1), 84–93.
- Mitchell, T. P., and J. M. Wallace (1992), The annual cycle in equatorial convection and sea surface temperature, *J. Clim.*, *5*(10), 1140–1156.
- Moore, A. M., and R. Kleeman (1999), The nonnormal nature of El Niño and intraseasonal variability, *J. Clim.*, *12*(10), 2965–2982.

- Neelin, J. D. (1991), The slow sea surface temperature mode and the fast-wave limit: Analytic theory for tropical interannual oscillations and experiments in a hybrid coupled model, *J. Atmos. Sci.*, *48*(4), 584–606.
- Neelin, J. D., and F.-F. Jin (1993), Modes of interannual tropical ocean–atmosphere interaction—A unified view. Part II: Analytical results in the weak-coupling limit, *J. Atmos. Sci.*, *50*(21), 3504–3522.
- Neelin, J. D., D. S. Battisti, A. C. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak (1998), ENSO theory, *J. Geophys. Res.*, *103*(C7), 14,261–14,290.
- Newman, M., P. D. Sardeshmukh, and C. Penland (2009), How important is air-sea coupling in ENSO and MJO evolution?, *J. Clim.*, *22*(11), 2958–2977.
- Newman, M., S. I. Shin, and M. Alexander (2011a), Natural variation in ENSO flavors, *Geophys. Res. Lett.*, *38*, L14705, doi:10.1029/2011GL047658.
- Newman, M., M. Alexander, and J. D. Scott (2011b), An empirical model of tropical ocean dynamics, *Clim. Dyn.*, *37*(9–10), 1823–1841.
- Newman, M., P. D. Sardeshmukh, C. R. Winkler, and J. S. Whitaker (2003), A study of subseasonal predictability, *Mon. Weather Rev.*, *131*(8), 1715–1732.
- Penland, C. (1996), A stochastic model of IndoPacific sea surface temperature anomalies, *Phys. D: Nonlinear Phenom.*, *98*(2–4), 534–558.
- Penland, C., and L. Matrosova (1994), A balance condition for stochastic numerical models with application to the El Niño–Southern Oscillation, *J. Clim.*, *7*(9), 1352–1372.
- Penland, C., and P. D. Sardeshmukh (1995), The optimal growth of tropical sea surface temperature anomalies, *J. Clim.*, *8*(8), 1999–2024.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003), Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, *J. Geophys. Res.*, *108*(D14), 4407, doi:10.1029/2002JD002670.
- Sardeshmukh, P. D., G. P. Compo, and C. Penland (2000), Changes of probability associated with El Niño, *J. Clim.*, *13*(24), 4268–4286.
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), *J. Clim.*, *21*(10), 2283–2296.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.*, *106*(D7), 7183–7192, doi:10.1029/2000JD900719.
- Vimont, D. J., M. Alexander, and M. Newman (2014), Optimal growth of central and east Pacific ENSO events, *Geophys. Res. Lett.*, *41*, 4027–4034, doi:10.1002/2014GL059997.
- von Storch, H., and F. W. Zwiers (1999) *Statistical Analysis in Climate Research*, p. 484, Univ. Press, Cambridge, U. K.
- Wittenberg, A. T. (2009), Are historical records sufficient to constrain ENSO simulations?, *Geophys. Res. Lett.*, *36*, L12702, doi:10.1029/2009GL038710.
- Wittenberg, A. T., A. Rosati, T. L. Delworth, G. A. Vecchi, and F. Zeng (2014), ENSO modulation: Is it decadal predictable?, *J. Clim.*, *27*(7), 2667–2681.
- Xue, Y., A. Leetmaa, M. Ji, A. Leetmaa, and M. Ji (2000), ENSO prediction with Markov models: The impact of sea level, *J. Clim.*, *13*(4), 849–871.
- Zhao, M., H. H. Hendon, O. Alves, G. Liu, and G. Wang (2016), Weakened eastern Pacific El Niño predictability in the early twenty-first century, *J. Clim.*, *29*(18), 6805–6822.